DOCUMENT RESUME

ED 086 952 CS 000 889

| | |
|---|---|
| AUTHOR | Rankin, Earl F. |
| TITLE | The Measurement of Reading Flexibility: Problems and Perspectives. Reading Information Series: Where Do We Go? |
| INSTITUTION | International Reading Association, Newark, Del. |
| SPONS AGENCY | National Inst. of Education (DHEW), Washington, D.C. |
| PUB DATE | 74 |
| NOTE | 62p. |
| AVAILABLE FROM | International Reading Association, 800 Barksdale Rd., Newark, Delaware 19711 ($1.50 nonmember, $1.00 member) |
| EDRS PRICE | MF-$0.65 HC-$3.29 |
| DESCRIPTORS | Readability; Reading; Reading Habits; *Reading Improvement; Reading Materials; *Reading Research; *Reading Skills; Reading Speed; *Reading Tests |
| IDENTIFIERS | *Reading Flexibility |

ABSTRACT

This study is concerned with the measurement of reading flexibility. The contents include: "Review of Literature," which discusses the different concepts of reading flexibility and presents several definitions followed by a summary of variables revealed in these definitions; "Synthesis of Strengths and Weaknesses," which looks at techniques of measuring reading flexibility, concepts of reading flexibility, rate measurement, readability of materials, and external flexibility; and "Recommendations," which makes recommendations for future research and development in the measurement of reading flexibility, based upon the findings of this paper. (WR)

ED 086952

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# THE MEASUREMENT OF READING FLEXIBILITY:

## Problems and Perspectives

Earl F. Rankin
University of Kentucky

eo+ira

Reading Information Series: WHERE DO WE GO?
1974

International Reading Association
800 Barksdale Road
Newark, Delaware 19711

CS 000 889

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

International Reading Association

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

The **Reading Information Series: Where Do We Go?** from Eric/Crier+IRA is designed to review the past, assess the present, and predict the future. This paper reflects the continued careful and thoughtful development of the series which was initiated by Dr. Richard A. Earle and the Eric/Crier staff at Indiana University. The titles currently available in the series are listed on the back cover of this volume.
Copies are available from International Reading Association, 800 Barksdale Road, Newark, Delaware 19711.

**James L. Laffey**
Director of Eric/Crier

**Billie S. Strunk**
Publications Editor

**Bruce Tone**
Director of Publications

The International Reading Association attempts, through its publications, to provide a forum for a wide spectrum of opinion on reading. This policy permits divergent viewpoints without assuming the endorsement of the Association.
The material in this publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to the International Reading Association for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either the International Reading Association or the National Institute of Education.

ISBN 0-87207-866-3

# Foreword

In teaching reading to young people in the elementary and secondary schools, one of my objectives was to help students adapt their rate and style of reading to their purpose and to the material. In working with both inservice and preservice classes, I made certain that the importance of this skill—*reading flexibility*, to use the term suggested by the author of this publication—was frequently presented. As an adult reader, and an observer of the reading habits of other adults, I am aware that flexibility has all but vanished from our repertoire of skills and that many of us are rather *inflexible* in our reading habits.

In *title*, Dr. Rankin presents the concept that flexibility involves the reader's reading skills, his psychological state, and the difficulty of the material he reads. Certainly, such a concept is consistent with the view that the act of reading is basically an aspect of the thinking process. Dr. Rankin reinforces the importance of the teacher's understanding of the learning processes and his realization that effective teaching of reading is premised upon knowledge of the learning style of each pupil.

The officers and members of the Board of Directors of the Association express their appreciation to Dr. Rankin for this contribution to the Association and to more effective learning to read by young people and adults throughout the world.

Millard H. Black, *President*
International Reading Association
1973-1974

# Preface

The writer wishes to acknowledge the contribution of two papers in helping him locate significant references and in suggesting valuable insights related to measurement problems. One paper is an unpublished manuscript entitled "Reading Flexibility: An Investigation" by Kathleen A. Jongsma (1971); and the other is "Assessment of Flexible Efficient Reading" by Phil L. Nacke (1971).

This monograph is an outgrowth of an earlier paper entitled "The Measurement of Reading Flexibility" printed in the *Occasional Papers in Reading* series under the auspices of the Reading Program of the Indiana University School of Education. It is, essentially, a revision of the original paper with substantive modifications and editing changes.

# Contents


Foreword 3

Preface 4

6 Introduction

7 Review of Literature
- Concepts of Reading Flexibility 7
- Techniques of Measuring Reading Flexibility 10

37 Synthesis of Strengths and Weaknesses
- Areas of Agreement and Positive Findings 37
- Critique 41

47 Recommendations
- Model of Reading Flexibility 47
- Implications of the Model 47
- Other Recommendations 51

57 References

62 Test References

# Introduction

This study is concerned with the measurement of reading flexibility. As such, it emphasizes primarily different measurement procedures and does not attempt to present a comprehensive investigation of the results of such procedures which have been utilized many times in the study of reading flexibility and its correlates. Since various concepts of reading flexibility influence techniques of measurement, it is necessary first to consider a few representative concepts as they are revealed in verbal definitions and a summary of variables indicated in these definitions, only some of which have been used in research on the measurement of different types of reading flexibility. Second, for reasons to be given, certain kinds of measurements or studies are excluded from this review. Third, published tests which purport to measure reading flexibility are described and evaluated, and then a variety of informal tests and measurement procedures found in the research literature, which might have implications for the construction of more adequate tests of reading flexibility, are reviewed.

Following the review of the literature on concepts of reading flexibility and techniques for their measurement, a summary of areas of agreement and positive findings is presented. This is followed by a critical evaluation of instruments and techniques of measurement which have been used in previous research. An attempt is made to point out both the strengths and weaknesses which characterize efforts to measure this important aspect of reading.

The concluding section of this study includes a proposed model for reading flexibility which suggests needed research and development on the measurement process. Finally, other recommendations are presented for research and development of more valid and useful measurements of reading flexibility in the years ahead.

# Review of Literature

## Concepts of Reading Flexibility

Many different concepts of reading flexibility have been found in research on this topic. Differences in concepts are reflected in verbal definitions by "authorities" using a number of different variables which have influenced the construction of measurement instruments. Several definitions are now presented, and these are followed by a summary of variables revealed in these definitions. Finally, some types of excluded measurement procedures and studies are specified which do not include measurements conforming to most concepts of reading flexibility.

### *Definitions*

The following quotations are probably representative of different points of view or degrees of emphasis about components of reading flexibility. Perhaps one of the earliest definitions was presented by Carrillo and Sheldon (1952):

> The mature reader is the adaptable, versatile reader; he should be able to adapt his rate of reading to the purpose with which he approaches the printed page, and to the difficulty level of the material. The goal is understanding at an adequate level.

A different emphasis was observed in this definition by Berg (1967):

> In general, the term refers to the activity a reader is engaged in when he sets up various patterns of thinking relative to his reading needs and then selects the skills that best accomplish this purpose. The term also implies that the reader can carry out the reading activity selected with an optimum of comprehension for the time expended.

A very broad concept was indicated in Stauffer's definition of flexibility as ". . . a high rate of efficiency in satisfactory attainment of the reader's purpose" (1962). McDonald (1963, 1965, 1967) in his

many writings on this subject has consistently rejected the notion that flexibility is the adjustment of rate best suited to purpose and reading material. He has emphasized the adjustment of *reading approaches* (i.e., perceptual and cognitive processes, reading skills, study techniques) as being necessary to gain an understanding of the author's meaning as dictated by the reader's purpose. He also has included the concept of a minimum expenditure of psychological and physiological effort within his concept of reading flexibility.

McCracken (1965) has made a distinction between *internal flexibility* (the adjustment of rate and approach within the sentences and paragraphs which make up an article) and *external flexibility* (similar adjustments between total passages). This is an important distinction because very little study has been given to the investigation of internal flexibility.

### *Summary of variables*

The previous definitions are in agreement that reading flexibility involves a relationship between one or more dependent variables involving some changes in reading behaviors and some one or more independent variables involving differences within the reader or within and/or between materials. Mere changes in reader behavior from one point in time to another are not indications of reading flexibility.

Several different variables have been indicated in these varying concepts. Independent variables include reader purpose and difficulty of material. Dependent variables include reading rate, reading approaches, minimum effort, maximum efficiency, purpose attainment, and optimum comprehension.

### *Excluded Studies*

It is evident that there is no such entity as reading flexibility. Rather, there are different types of reading flexibilities. The following review of techniques of measurement reinforces this conclusion. However, there are types of studies sometimes considered as reading flexibility investigations, which do not involve the measurement of reading

flexibility as indicated by most definitions. One of these is an investigation which manipulates differences between individuals as independent or dependent variables. For example, if two equated groups read two passages written at different levels of difficulty and displayed significant mean differences in rate; this would not be a measurement of reading flexibility. Such a study would suggest hypotheses which could be included in the measurement of reading flexibility, but all definitions agree that intraindividual differences in behavior as a function of some one or more independent variables is an essential ingredient in the concept of reading flexibility. These types of interindividual studies will be excluded from this paper except insofar as they suggest important ideas for the measurement of reading flexibility.

Another type of study to be excluded from this survey is the measurement of the relationship between rate and comprehension. The writer considers both rate and comprehension as dependent variables which result from changes in reading approaches as a function of the manipulation of some one or more independent variables. Therefore, such studies do not measure reading flexibility as the term is used by this writer.

As previously stated, studies of mere changes in reading behavior over time unrelated to some independent variable, are not included within the concept of reading flexibility.

Finally, reading flexibility is essentially a positive concept. It is a desired outcome of learning. It does not include any change in reader behavior as a function of a change in an independent variable. Therefore, the study of the relationship between interest-appeal of materials and reading rate which might indicate a correlation (either positive or negative) between the two variables does not measure reading flexibility. Certainly, the measurement of the relationship between material difficulty and comprehension, which shows that comprehension declines as material difficulty increases, would give no information about reading flexibility. The measurement of reading flexibility necessarily entails a study of the ability of readers to adjust their behavior under two or more conditions so as to accomplish their reading purpose(s). Such changes in behavior are reflections of desirable changes in reading approaches. A further discussion of this conceptualization of reading flexibility is presented in the chapter on the "Model of Reading Flexibility."

## Techniques of Measuring Reading Flexibility

Various published tests constitute operational definitions of reading flexibility which have influenced the outcome of many research studies. There are only a few published tests of this skill, and these tests are, in the writer's opinion, rather primitive from the standpoint of adequate standards of measurement. These are now described and evaluated. Following the review of published tests, a number of unpublished tests and measurement procedures used in research on reading flexibility are considered.

### *Published Tests*

*Test of Reading Flexibility.* This test was devised by Spache and Berg in 1958 for use by college students and adults (Spache, 1956). It was published in a book entitled *Faster Reading for Business*, now out of print. The writer believes this was the first published test of reading flexibility. The test attempts to measure flexibility by studying the effects of variations in reading the same article three times for three different purposes. The reading passage is a 2,800 word article about forecasting for business. On the first reading (skimming), the reader is given a three minute time limit to read the article and told that he will be asked to answer ten questions on main points without looking back at the article. The second reading (scanning) requires the reader to read ten questions in advance and find the answers by referring to the selection. The timing on this reading includes the time it takes to answer all questions. The third reading (reading for a thorough understanding) is read without time limits with instructions to read in order to answer twenty detailed questions without looking back at the selection. Timing is based on reading time only. The questions cover facts, inferences, and conclusions.

The Test of Reading Flexibility attempts to measure the relationship between *purpose* as an independent variable and *rate* measured as time spent. Comprehension is measured as attainment of the assigned purpose and must be adequate, as indicated by norms, in order for rate measurement to be interpretable. A very desirable feature of this test is that material and reader variables such as difficulty, background information, and interest appeal are held constant while purpose is

(presumably) systematically varied by specific and clearly stated reading objectives. Of course, the extent to which individual readers accept or reject the assigned purposes is not known. The reading passage is longer than passages used in many other tests of reading flexibility. Unlike a subsequently developed test battery which attempts to measure the results of skimming with only three questions and the results of scanning with only one question (i.e., Reading Versatility Tests), this test provides ten questions to measure comprehension on both skills. Twenty questions are used to measure thorough comprehension, and a test of this length might have suitable reliability. All test questions are multiple-choice with four alternatives.

Separate norms are given on a five point scale ranging from *poor* to *excellent* for both rate and comprehension. The book provides an extensive discussion for the interpretation of test results. No attempt was made by the authors to establish criteria for the interpretation of differences in rate, as such. As will be noted later, this is a distinct advantage. Instead, the reader can interpret his rate on scanning and thorough reading individually on a normative basis, provided that his comprehension was satisfactory according to the norms.

Unfortunately, no information was provided by the authors regarding the size or characteristics of the normative sample. Judging from Spache (1956), the normative data were gathered informally by loaning test copies for trial use in exchange for accumulation of test results for various groups. No information was given in the book about the reliability or validity of either rate or comprehension scores. In fairness to the authors, it must be admitted that this book was hardly an appropriate place for technical information. However, the writer is informed by Dr. Berg that the test was constructed informally without obtaining this kind of technical data. It should be noted that the subject matter of the reading passage is more appropriate for adults in business than for many college students.

The Test of Reading Flexibility, although perhaps the first published test of reading flexibility and now out of print, has many desirable features which might well be emulated by future constructors of flexibility tests. Due to the excellent control over many variables, other than purpose, which might influence changes in rate and comprehension, the results of this test are more easily interpretable than the results of several tests which have been published subsequently.

*Flexibility of Reading Test.* Constructed by Braam and Sheldon, two of the three forms of this test were published in a book for college students entitled *Developing Efficient Reading* (1959). The third form has not been published. The test contains five passages representing five different types of material: narrative, literature, science, history, and psychology. The articles range in length from 750-900 words. An attempt is made to hold purpose constant for all selections by instructing the reader to read as quickly as he can and still understand the general content of the selection. The reader is timed on each passage to obtain a rate measured in words per minute. No time limits are used for reading the materials. After reading each article, the reader attempts to answer ten true or false questions. Strangely, the reader is not told whether to refer to the selections in answering questions. Presumably, he is expected to answer these questions without referring to the text. The degree of flexibility is measured by the amount of difference between the slowest and the fastest rates on the five passages for a given student. Comprehension scores are expressed as percentages.

As Braam (1963) has noted, there were several problems of control in the construction of this test. He rightly pointed out that purpose for reading could not really be controlled by test instructions. He noted that the Dale-Chall readability formula measurements for each passage showed the passage difficulties to range from grades nine-ten through grade sixteen. The directions for the test inform the student that he is to read passages covering different subjects which are of varying levels of difficulty and familiarity.

In the light of these confounded independent variables, it is difficult to interpret the meaning of the test results. Differences in rate and/or comprehension may be due to any one or more of these variables for a particular student. A table was provided for interpreting flexibility (i.e., difference in rate between the slowest and fastest passage) on a seven point scale ranging from very poor to outstanding. However, the basis for these normative categories was not explained in the book. More importantly, the use of difference scores to measure flexibility in this and other tests raises two very fundamental questions.

First, as Thorndike and Hagen (1961) have shown, when a difference is taken between two test scores, the reliability of the difference is usually much lower than the reliability of the two tests

upon which the difference was based. This reduction in reliability is due to the fact that, in the subtraction process, whatever factors are common to both measures are cancelled out and the difference score contains only those factors which are specific to both tests plus the errors of measurement of both tests put together. It can be shown that this reduction in reliability of the difference score increases as the correlation between the two individual measurements increases. Theoretically, if the correlation between the two tests were perfect, the difference scores would reflect only errors of measurement; that is, the reliability would be zero (Thorndike and Hagen, 1961). Since reading rate scores tend to be highly correlated despite differences in difficulty and nature of materials (Bloomers and Lindquist, 1944; Carlson, 1951), it is unlikely that difference scores in reading rate will have satisfactory reliability for the measurement of reading flexibility, even if the reliabilities of each individual rate test are substantial. Hence, on statistical grounds alone, it is highly improbable that the flexibility measurements in this test have suitable reliability and, therefore, validity. Second, even if the difference scores on this or other tests had satisfactory reliability, the question still remains as to a possible interaction effect between difference scores and status on the lowest rate measurement involved in the difference, due to the measurement procedure itself. This procedure involves subtracting the difference in rate between the slowest and fastest rate. Unless it can be shown that there is no such test-induced interaction, a person whose lowest rate score is very fast might not have the same chance to make a high flexibility score as a person whose lowest rate score is rather slow. If this were the case, difference scores established at different levels of speed would not be comparable measurements.

The interpretation of comprehension scores in this test is not clear. Obviously, ten true or false questions are not a reliable measurement of comprehension. No norms are provided for these tests. The reader is cautioned that a rapid rate (?) accompanied by a comprehension score of below 80 percent and a feeling of lack of understanding " . . . would, of course, not be a valid basis for comment" (Braam and Sheldon, 1959). He is also told that a comprehension score of 70 percent or below indicates a need to read more slowly. Apparently these percentage criteria were arrived at in a completely arbitrary manner. Their chief function is probably to keep the rate scores "honest." They do not constitute an adequately measured dependent

variable of comprehension purpose attainment in the test of flexibility.

According to Braam (1963), comparability of test forms was accomplished by taking two selections for each of the five passages from a common source. No evidence was given for the effectiveness of this procedure. Also, no evidence was given regarding comparability of test questions either within or between forms. Order effects in reading a sequence of passages were apparently not considered. The authors provided no evidence of reliability or validity of measurements for rate, comprehension, or flexibility.

However, Berger (1969) studied the interform reliability of the rate scores and modified multiple-choice comprehension questions for the three forms of the test. The true or false comprehension items were apparently modified to reduce the effects of guessing on these scores. The rate reliability coefficients (based on seventy college students) between forms for the same subject areas ranged from .56 to .83. These reliability figures were quite low for individual or group use. Often, the interform rate correlations were higher *between* subject than *within* subject areas. The reliability coefficients for comprehension items within subject areas ranged from −.35 to .30. These figures were, of course, quite low. Many correlations did not even attain statistical significance at the .05 level. Again, many of the between-area correlations were higher than the within-area correlations. If Berger's multiple-choice questions were an improvement over the true or false questions in the published test, the reliability (and, hence, the validity) of the published comprehension tests must be low indeed. Berger (1969) found high rate reliabilities (but low comprehension reliabilities) for the test as a whole. However, the relevance of these findings is not clear. The essential reliability of a test of flexibility depends upon the reliability of rate and comprehension scores for the subtests, not the total test.

It must be concluded that the Flexibility of Reading Test is a highly defective instrument lacking in most of the accepted technical prerequisites for good test construction. It might be assumed that this test, like the previously described test by Spache and Berg, was published as part of a workbook and, therefore, is not subject to rigorous conventional technical criteria as applied to standardized tests. However, the test was used by Braam as a measuring instrument in a scientific study. An attempt by the writer to obtain technical data about the test from the authors proved unsuccessful.

*Reading Versatility Tests.* These tests were first published in 1962 by McDonald et al. and were revised in 1968. This was the first published battery of tests of reading flexibility to be used by readers of various grade levels and to provide four equated forms at each reading level. Whether the term *reading level* is to be interpreted as reading achievement level or as grade level is not clear in the *Manual of Directions* (McDonald, 1968). The three levels are *Basic* (levels 5-8), *Intermediate* (levels 8-12), and *Advanced* (levels 12-college).

Each test contains four selections, each differing in difficulty, style, and theme. Each of these four passages is supposed to be read for a different assigned reading purpose which requires different approaches. McDonald speaks of such variables as skills, ways of thinking, and psychological set as *approaches.* According to the manual, the first selection is a fiction passage which is to be read rapidly but with attention directed to important facts, main ideas, and details. The second selection is a nonfiction passage which is to be read carefully with attention to details, main ideas, and implications. The reader is supposed to skim the third passage looking for main ideas. The fourth passage is to be scanned in order to locate the answer to one question provided in advance. All questions, with the exception of the scanning task, are answered without looking back at the passage. There are no time limits. Reading time before answering questions is converted to a words-per-minute score. Ratios of reading rates for various selections are used to indicate efficiency in varying reading approaches for different purposes (i.e., flexibility). Selections one and two are followed by ten questions, the third selection (to be skimmed) is followed by three questions, and the fourth selection (to be scanned) requires the student to answer only one question.

At first glance, the Reading Versatility Tests appear to be an impressive and ambitious attempt to provide a new and better measurement of reading flexibility for readers of many ages. It has all of the advantages of multiple forms and publication of individual tests free from the confines of a book. This was not the case with the two previously described tests. However, this series of tests contains some serious flaws.

No norms are provided in the manual for rate measurements on these test materials. Instead, the reader is referred to the Educational Developmental Reading Laboratory norms established for the Reading Eye Camera. These norms were established on different materials of short duration with statements of purpose different from those used

in the Reading Versatility Tests. Furthermore, these norms were limited to readers at each grade level who, while attached to a camera, attained at least a 70 percent score on comprehension tests. Consequently, these results are probably not representative of readers in general. Some impressive interform reliability coefficients for rates are given in the manual for selections one and two ranging from .82 to .90. Detailed reliability coefficients are, for some reason, not given for all forms of the test at each level. These reliability coefficients in the manual may, or may not, be representative of reliability coefficients for all test forms of the Reading Versatility Tests. The rate reliability measurements for selections three and four combined are quite low with correlations ranging from .51 to .70. Had these two parts not been combined, the rate reliabilities would undoubtedly have been lower. It is, of course, the reliabilities of *individual* parts of the total test which are particularly relevant to the measurement of reading flexibility.

Differences between rates on various subtests are converted to rate ratios by dividing the rate obtained on a particular passage read for one purpose by the rate obtained on another passage which was read for another purpose. Criteria were provided in the manual for the interpretation of the efficiency of these ratios, but no basis was given for the manner in which these criteria were determined. It is significant that no reliability figures at all were given for these rate ratios in the manual. If other investigators have obtained this information, the writer has been unable to locate such studies. It may be that McDonald et al. simply assumed that if the individual rate measures were sufficiently reliable, consequently it could be logically deduced that the ratios based on these rate measurements would be reliable also. Unfortunately, this is not necessarily true. Rate ratios based upon differences between rates may have the same deficiencies in reliability as difference scores, themselves. This problem is the same as described in the section on the Flexibility of Reading Test by Braam and Sheldon. Even with adequate individual rate reliabilities (which all four subtests in each total test of this battery do not have) the reliability of a difference score on correlated measurements must be low. The following quotation by Gulliksen (1950) points out the similarities between reliability deficiencies of difference scores and ratio scores which are simply another way of expressing differences:

> When the accomplishment quotient (AQ) was introduced, Kelly pointed out that the problem involved was to obtain reliable measures of each variable.

Clearly these measures would be correlated so that the accomplishment quotient might reflect only errors of measurement.

Now if the criteria which McDonald et al. used for efficient rate ratios were based on minimal rate differences sufficiently large to exceed the standard error of the measurement for particular differences in scores involved in computing rate ratios, this might be an improvement in test construction. However, since they used the same criteria for interpreting rate ratios on all test forms at all levels, it is unlikely that this technique was used. The use of the same criteria for interpreting the efficiency of different rate ratios based on different levels of the distribution of different tests is questionable. As Cronbach (1941) pointed out, "... ratio scores within the same population and even ratios which are equal in size, may not have the same reliability. The standard error of the measurement increases as the denominator decreases."

In addition to the problem of the reliability of these rate ratios, like the difference scores there may be a test induced statistical interaction between the status of the base rate for the difference and the ratio obtained. Until this issue is resolved, it is not meaningful to talk about slow readers with a high rate ratio versus fast readers with a low rate ratio, etc. A desirable feature of the test was an attempt to reduce the effect of previous information on comprehension test performance for selections one and two by using reading related items. Questions which could be answered correctly by 40 percent or more of the students who had not read a given test passage were discarded or rewritten. This procedure was apparently successful, because, despite the short length of the comprehension tests for selections one and two, concurrent validity coefficients between these subtests obtained for several groups and the Diagnostic Reading Test: Survey Section; the Level of Comprehension subtest of the Cooperative Reading Tests: Reading Comprehension; and the Davis Reading Test were quite substantial with correlations ranging from .67 to .89.

Despite the impressive concurrent validity of the comprehension tests for selections one and two, their interpretation for an individual student is subject to question because no norms were provided. Instead, a 60 percent criterion was recommended as a minimum score on both selections one and two if the rate scores were to be considered meaningful. If it is assumed that selections one and two were each written at substantially different levels of readability, and if the test for each passage was a valid measure for the comprehension of

that particular passage, how would it be possible to use the same 60 percent criterion for the comprehension of both passages?

There were also no norms or reliability coefficients provided for the skimming and scanning comprehension tests in the manual. The manual instructs the reader to disregard the rate scores for skimming if the student missed more than one question (out of three) and on scanning if he missed the one (and only) question on that test. No evidence was given for the manner in which these criteria were established. On such a small number of questions and by using these apparently arbitrary criteria, how can the student be informed that "He may be trying to use skimming and scanning techniques but lacks the requisite skills" (McDonald et al.)? Also as Maxwell (1969) has indicated, adequate reliability of comprehension on skimming and scanning cannot be obtained with only a few items.

The manual claims that comparability of forms was obtained by using articles in each form for each part which were comparable in difficulty, content, interest appeal, and style. Curiously, however, no evidence (other than a few selected interform reliability coefficients) was provided in support of these claims.

All of McDonald's writings on reading flexibility have placed a great deal of emphasis upon adjustment of reading approaches rather than reading rates as fundamental to his concept of reading flexibility. In the writer's opinion, however, it is precisely this factor of approach which the Reading Versatility Tests do not really measure adequately. Although space does not permit documentation of this point, McDonald's statements of purpose are vaguely stated and might induce a variety of approaches in different readers. Also, the comprehension questions do not necessarily reveal the approach used. As Nacke (1971) has indicated, there are different skimming strategies which are not measured by answering test questions. There also may be a variety of approaches to finding a main idea or remembering details which are not revealed by answering test questions. Due to the many confounded variables which influence performance on the Reading Versatility Tests (difficulty, style, content, interest appeal, purpose), different individuals might be induced to use a variety of approaches on the same subtest, and this would not necessarily be reflected in their answers to test questions. The manual does claim that evidence from eye-movement photographs were consistent with the assumption that differential approaches were being used. But little empirical

evidence was cited in the test manual (other than a reference to a research paper by Welch, 1964) to support this contention. In any case, the question still remains about the adequacy of eye-movement photographs to reflect the many subtleties of cognitive and perceptual functions involved in reading approaches. No introspective evidence known to the writer has been gathered on these tests to obtain evidence about the reader's acceptance of assigned purposes or his perception of his approaches used to attain these purposes. Even if comprehension test results were conclusive evidence of approach used, the tests do not make provisions for analysis of test results on selections one and two to determine purpose attainment such as main ideas, details, and inferences. Rate, rather than approach, is the only carefully measured dependent variable in this test, but even this measurement is of doubtful validity expressed in ratio form. As the writer has previously indicated, the causal variables which might influence test performance are so confounded, that test results are difficult to interpret. McDonald et al. (1968) maintained that only 5 percent of high school and college students are flexible as measured by his test. Would greater differences in readability between selections one and two produce greater rate ratios? Do confounded influences of style, background information, or interest appeal operate, perhaps, to reduce the effects of differences in readability upon rate ratios? Would different criteria for the interpretation of rate ratios based on adequate empirical evidence show that perhaps people are more flexible after all? More fundamentally, are rate ratios valid measures of differences in approach used in reading for different purposes? Would more reliable measures of comprehension in skimming and scanning, together with more realistic comprehension criteria for eliminating consideration of these rate scores for individual students because of inadequate comprehension, reveal a higher degree of reading flexibility?

These and many other questions remain unanswered. Given the many inadequacies of this test, results of studies which have been based on the use of the Reading Versatility Tests should be regarded with caution. It is unfortunate that so many often quoted findings and conclusions about the nature and extent of reading flexibility have been based upon measures provided by the Reading Versatility Tests.

*Reading Test.* Published in 1970, this test was constructed by Raygor for use by college-bound high school juniors and seniors and

also college freshmen and sophomores. Two forms of the test have been published. Reading flexibility is only one of several skills measured by this test, and the comments in this paper are limited to this part of the total test and the test for skimming and scanning. Flexibility is measured by comparison of reading rates on two articles written at different levels of difficulty and written about different topics. The first article resembles easy recreational material and is, according to the manual (Raygor 1970), written at the level of difficulty of most newspapers, magazines, and novels. The second article is similar to more difficult study materials found in college textbooks. Reliability of rate scores for these passages was not given in the manual. Rate, in words per minute, is measured in each article by the number of words read during the first three minutes of reading time. The reader is given a total of five minutes to complete each article. The first passage is roughly 1,800 words in length; the second, about 1,600 words.

Each passage is followed by ten multiple-choice questions with four alternatives. These questions measure only the retention of factual content. A ten minute time limit is given for answering test questions. There is no specifically assigned purpose for reading the two articles involved in the measurement of flexibility. The purpose is established by describing the general nature of the material in the two passages and informing the readers that they will be required to answer ten test questions on each passage which will determine how well they have understood the material. Readers are told the time limits for reading and answering questions before they start each part of the test.

Another part of the test is designed to measure skimming and scanning. Students are told they will have to obtain information quickly without actually reading all the material. They are given thirty test questions in advance which they may attempt to answer by referring to various kinds of materials (indexes, charts, bibliographies, and textbook excerpts), all of which are printed on blue pages in contrast to the white pages in the rest of the text. The reader is told that he will have ten minutes to complete all thirty items.

Unlike all previous tests of reading flexibility, excellent norms were provided for this test based on a large standardization sample. Separate norms were obtained for college-bound juniors and seniors in high school, four-year college freshmen and sophomores, and two-year college students. Percentile, standard scores, and stanine norms were

provided for rates on both easy and difficult passages, rate differences (i.e., flexibility), combined retention tests on both passages, and comprehension scores for skimming and scanning.

In several respects, this test represents a distinct improvement over other published tests. The normative data are comprehensive and are based on a large sample of subjects representing a large geographic area. The skimming and scanning test is an improvement over previous tests. It utilizes many different kinds of materials and a large number of test items. In the writer's opinion, this latter test might better be called a "scanning" test since the questions tend to direct the readers' search for the answers to factual questions rather than general ideas. A rather important feature of this test is the manner in which purpose is induced. Instead of giving specific assigned purposes, the reader is given a brief description about the nature and difficulty of each passage involved in the measurement of reading flexibility, and in the light of this general information he is allowed to make his own choice regarding rate and approach. The ability to choose one's own purpose in the light of relevant information is an important dimension of reading flexibility.

There are, however, some deficiencies in this test which should be considered. The rate measurements are based on only the first three minutes of reading. Traxler (1938) found that measuring reading rate during a short time (i.e., from one to five minutes) resulted in unreliable measurements. Although the comprehension items for the flexibility test are limited to factual items, the student is not informed of this fact in advance. Therefore, many students might be reading for other more study-type purposes, particularly on the textbook selection and might consequently make a low score on factual items. Strangely, norms were provided only for combined tests for both passages involved in the measurement of reading flexibility, so the meaningfulness of the individual rate score from each passage cannot be interpreted in arriving at the measure of flexibility. The announcement of time limits in advance of testing for each part of the test might have differential effects on individuals who have or do not have watches or groups who have or do not have wall clocks. Of course, the resulting rate differences for individual readers between the two test selections used in the measurement of flexibility may be due to a number of confounded variables such as purpose, article difficulty, interest appeal, background information, etc. More fundamentally, the

previous discussions of reliability problems of difference scores apply to this test and cast serious doubt as to the reliability and validity of these difference scores and the meaning of norms based upon them. Unless the rate difference scores on this test have a very high reliability combined with a low correlation between tests, small differences in rate are not reliable, and norms based on them are essentially meaningless. According to the manual, the correlation between the two rate scores on this test is .70. From Table 7.8 in Thorndike and Hagen (1961) it can be deduced that, if the average reliability of these two rate scores were .80, the reliability of the rate difference scores would be .33. In the light of this observation, it follows that the possible validity of these flexibility measurements is limited and the norms based upon small rate differences are meaningless.

The manual (1970) gives twenty Kuder-Richardson reliability coefficients for several parts of the test and the total Reading Test. The authors indicated that these figures are spuriously high on the skimming and scanning comprehension results because of the speed factor. No separate reliability figures were given for comprehension scores on the easy and difficult passages involved in the measurement of flexibility. Even on the combined comprehension tests for these two passages, the reliability is only .65. It is understandable, therefore, that the individual comprehension test results for these two different passages are ignored. No reliability information was given for either rate or rate differences in the manual. High content validity is claimed for the test. It is also claimed that information about comprehension items which could be guessed without reading the passages is used in constructing the test. Precisely how this information was used and the results of such use are not revealed.

Correlations were computed between the total score and various subtest scores of the Reading Test and the total and subtest scores of the Nelson-Denny Reading Test. These results might be of interest from a scientific point of view, but the Nelson-Denny Reading Test is really not a meaningful criterion for the concurrent validity of the rate scores, flexibility scores, or skimming-scanning scores of the Reading Test.

On the whole, the Reading Test includes some desirable features as a measure of reading flexibility not possessed by other tests. Unfortunately, like the previously described tests, flexibility is

measured by rate differences of unknown and probably very low reliability. It is strange that not a single one of the test authors who has attempted to measure flexibility by difference or ratio scores has been aware of this problem in test construction. If these authors were familiar with this measurement difficulty, it is not reflected in their tests or test manuals.

*Experimental measurements*

This section is devoted to consideration of a number of different measurement procedures which provide a variety of operational definitions of reading flexibility. Some of these procedures are unpublished tests designed, like published tests, to measure individual differences in reading flexibility. Other investigations involve the experimental manipulation of one or more independent variables in order to study their effects on changes involving various reader behavior variables which are treated as dependent variables in these investigations. Although it is not usually the purpose of these studies to measure individual differences in reading flexibility, they often do provide information which might be used in the construction of such tests. Some writers (e.g., Jongsma, 1971) have considered these latter types of studies as investigations of correlates of reading flexibility. However, in terms of the prevailing conception of reading flexibility indicated in both definitions and published measurements as consisting of certain changes in reader behaviors (dependent variables) brought about as a function of changes in or between materials or within individuals (independent variables), these studies are considered as research on variables involved in the measurement of reading flexibility. This section is organized in terms of different categories of independent variables: material variables and reader variables. Examples of studies within each category are given. They use different categories of dependent variables or different ways of measuring the relationship between these two kinds of variables. Other studies which were essentially replications of the same measurement procedures are not described. A few pieces of research are considered which, although they do not contain measures of reading flexibility as such, still shed some light on measurement problems.

*Independent variables: materials.* A number of investigations have related material characteristics, usually difficulty or content, to one or more aspects of reader behavior. Until the study by Rankin and Hess (1970) was conducted, all previous studies in this category measured differences in reader behavior in relation to differences between articles. These types of studies measured what McCracken (1965) has called "external flexibility."

Several early studies (Flesch, 1949; Brown, 1952) showed that rate and comprehension changed in relation to the readability of materials, but the studies left many variables uncontrolled. Letson (1959) conducted an important study which compared the relative effects of material difficulty and reader purpose upon rate and comprehension. Using college freshmen, Letson had subjects read an easy and a difficult article with a common assigned purpose. They were to read as rapidly as possible and still understand the material sufficiently to answer questions afterwards without looking back. The students also read two equally difficult articles for two different purposes: 1) to read as rapidly as possible for the story, and 2) to read for complete mastery of ideas and details. All passages exceeded 2,500 words and were said to be comparable with respect to subject matter. All articles were followed by comparable comprehension tests with respect to difficulty and other technical criteria. Letson neglected to give the readability levels of these articles and to describe the comprehension tests. Also he used a five minute reading time limit which undoubtedly kept many students from completing the passages. His findings, although not subject to statistical tests, indicated that differences in difficulty have considerably more influence on reading rate than differences in purpose. He even noted that instructions to read for mastery caused some students to read faster rather than slower. These findings have implications for future study of reading flexibility, and his design suggests a suitable way of measuring and comparing the effects of two different independent variables upon reading performance. He also made a separate analysis of data for those subjects whose rates were different than expected in comparing reading performance on both types of materials and for both purposes. He called this "negative flexibility." This is a phenomenom that is in need of further study. His potentially important findings must be considered as suggestive, since there was no information given about the precise nature of several important control variables and no statistical tests of significance were made.

A study similar to Letson's was carried out by Levin (1968) using bright ninth grade girls and a slightly different comparison of purposes. This study will not be described, for it contributes nothing new in the way of suggestions for measurement procedures. However, Levin's investigation might be consulted, because it is characterized by a much greater degree of scientific sophistication. Readability figures were given for all articles, reliability coefficients were given for tests, and tests of significance were made in drawing conclusions.

It should be noted that comprehension is only an incidental factor in the past two studies mentioned. Comprehension was used chiefly to induce the proper mental set for readers and to make the rate scores meaningful. In the Levin study, it could have been used as a measure of differentiated comprehension purpose attainment, but the same factual types of questions were asked for all articles. In the next section, it is pointed out that there has been some confusion about the function of comprehension measurements in reading flexibility studies.

The claim of both Letson and Levin to have constructed comparable tests of equivalent raw score difficulty to measure the comprehension both of materials of unequal difficulty and possibly of purposes of unequal difficulty, raises the question about their validity. Clearly, the difficulty of a test must reflect a difficulty of the reading passage itself or the intellectual processes involved in reading for different purposes. A solution to this problem would, of course, be to construct valid tests for a given purpose or material and to convert individual scores to standard scores for purposes of comparison. This has not been done often in the literature on reading flexibility.

Pitcher (1953) conducted an interesting study of the interacting effects of readability and type of material upon rate. His results were limited to good college level readers who attained scores of 70 percent or above on three sets of ten item comprehension questions. This manner of selecting subjects suggests a technique for measuring the influence of one or more independent variables upon any dependent reader variable with comprehension held relatively constant. Pitcher used three types of material—familiar, abstract, and technical—each of which included articles written at three levels of readability as indicated by the Flesch formula. He found highly significant differences between rates within a given type category for articles written at different levels of readability. However, passages with equal readability ratings did not produce equal rates across type-content areas.

This study suggests the importance of controlling type of material when studying the effects of readability differences upon reading rate or other dependent variables. Several studies have failed to do this.

A potentially useful design for experimental purposes was used by Nicholaw (1968) in constructing a test to measure the interacting effects of readability, subject matter, and purpose upon rate and comprehension for sixth grade students. Nicholaw used three subject matter materials: literature, science, and social studies. Each subject area contained articles written on fourth, sixth, and eighth grade levels of difficulty. Two purposes were assigned for students to attain in reading a given passage: understanding main ideas and significant details. Total tests consisted of eighteen subtests based on passages 400-500 words in length.

No empirical evidence was provided for reliability, validity, or comparability of test questions. Nicholaw's test is probably useful only for experimental studies of group differences. Some modification of this design might be used in future test construction, although the time involved in taking such a test might be prohibitive for practical purposes.

Differences in eye movements have been used often as dependent variables in relation to differences in the difficulty of materials. Early studies by Judd and Buswell (1922) and Walker (1938) demonstrated such a relationship, but no statistical tests were used. A more recent and more sophisticated study carried out by Taylor et al. (1960) is used as an example of this type of measurement. Taylor used eighth grade students of average reading ability who had demonstrated the ability to read with adequate comprehension materials designed for the Reading Eye Camera. His subjects read different materials while their eye movements were being recorded. These materials, from the Reading Eye test file, were written at three grade levels—grades four through six, junior high, and high school/college. He obtained significant differences by use of an analysis of variance test between number of fixations, duration of fixations, rate, and comprehension for junior high versus high school/college materials. An important finding of this study, with implications for measurement, was that no differences in eye movements were obtained with these sixth grade students between fourth and sixth grade materials or between fourth and junior high materials. The interpretation given by Taylor was that the difficulty of content does not significantly affect habits of reading performance when the material is at or below the reader's grade level.

It should be noted that these findings were based on different content which may have varied in interest appeal and background knowledge for different readers. In any case, the problem of the difficulty level of materials used in measuring reading flexibility needs thorough and systematic study under carefully controlled conditions.

These previously described studies on the effects of material difficulty upon various reader behaviors have involved comparisons between materials. McCracken's concept (1965) of "internal flexibility" involving changes in approach and rate within a passage was not used in measurement until recently. Humphrey (1957) had studied variations in rate per minute to minute within a 7,000 word article read by college students, but his technique for obtaining such measurements was not explained. In any case, since he did not relate these changes in rate to any independent variable, the writer does not consider this to be a measure of internal reading flexibility. A possible reason for the delay in the development of a technique for measuring intra-article changes in reader behavior is found in a study by McDonald (1960). He determined that both reading rate and comprehension of college students with high anxiety, as measured by a personality test, were impaired by testing procedures involving periodic interruptions. He concluded: "Timing procedures which produce periodic interruptions during the reading process should be avoided."

In their 1970 study, Rankin and Hess found that periodic interruptions did not impair the reading of college students of high anxiety, as measured by the SA-S Senior Scales when subjects were given a practice period to help them to adapt to these testing conditions. Using material in the Diagnostic Reading Test: Survey Section and the twenty test items in this test, Rankin and Hess had students read for the purpose of reading as rapidly as possible with understanding. The subjects consisted of members of reading improvement classes who were selected for these classes on the basis of scores falling below the 33rd percentile on the Cooperative English Tests: Reading Comprehension. Subjects were instructed to underline the word they were reading when the signal "mark" was given every fifteen seconds. Previously, the readability of every successive 100 word passage had been determined by a comparable group of subjects using the cloze procedure on these materials. Rate measurements were computed by determining the number of fifteen second interval markings used for each successive 100 word passage. A flexibility coefficient was computed by correlating the rate scores with the

readability scores for each passage. This procedure produces negative correlations for students who tend to slow down for more difficult passages and to speed up for easier passages. In the 1970 study, Rankin and Hess obtained a correlation for the total group by using, as a rate measurement, the average rate for the group on each 100 word segment and correlating this distribution of average rates with the distribution of cloze readability scores. Measurements taken before and after a one-semester reading course yielded correlations of −.34 before training and −.48 after training. The latter was significant at the .05 level. A pilot study showed that the average comprehension score was approximately 75 percent on these materials under these testing conditions.

Another study, based on the analysis of individual flexibility coefficients for each of the subjects used in the previous investigation both before and after training, was carried out by Rankin (1970-1971). This study found a wide distribution of flexibility measurements even among these poor readers. Since several studies have shown that better readers tend to be more flexible than poorer readers, Rankin's results call into question McDonald's previously mentioned observations about the lack of reading flexibility among readers of all ages.

This recently developed technique for measuring internal (intra-article) reading flexibility has a number of desirable features. It does not involve the use of difference scores with their reliability problems. Flexibility coefficients can be interpreted in the same manner as all correlation coefficients. Cloze test readability measurements are valid measurements of readability for the particular population of subjects being studied while reading these specific passages. All factors affecting the difficulty of each passage for a particular group of readers should be reflected in cloze readability measurements. It should be noted, however, that this recently devised measurement technique is strictly an experimental procedure. Until techniques are devised for machine scoring and computer conversion of fifteen interval markings to the "number of 15 second intervals used per 100 words," the time and effort involved to do this work by hand is prohibitive. At present nothing is known about the reliability of these measurements or their relationship to "external flexibility." It also seems likely that the magnitude of the correlations obtained by this technique would be influenced by the range of both readability and rate measurements.

*Independent variables: reader characteristics.* One of the most important reader characteristics used as an independent variable in studies of reading flexibility has been purpose. First, studies are considered which attempt to control purpose by giving specific directions to the reader. Following these studies, a few methods are described which give the reader more freedom in choosing his own purpose.

A study by Walker (1938) found differences in the eye movements of superior college readers among the following assigned purposes on passages of equivalent difficulty to be read for general idea, details, thorough knowledge, and answering a specific assigned question. Although no statistical tests of these differences were made, the study does suggest that assigned purposes may be accepted, at least by superior (above 90th percentile) college readers. In general, however, the question of acceptance of assigned purposes in reading flexibility tests is subject to question unless some evidence is produced which indicates such acceptance. Previously described investigations by Letson (1959), Levin (1968), and Nicholaw (1968) have all attempted to measure the effects of assigned purposes on various reader behaviors such as rate and comprehension.

Shores (1960) used a different method of studying the effects of assigned purposes on readers. Using adult level science materials for both sixth grade students and college students, he used two assigned purposes: read for the main idea and read to remember ideas in sequence. Unlike most studies of purpose, Shores attempted to determine differences in reading approaches through the analysis of written introspective reports, following each reading, concerning the manner in which each reader thought he had read the material and how each student thought an ideal reader would have read the article. The results of this investigation are not of significance from the standpoint of this paper, but the method used was important in possibly influencing later development of reading flexibility tests by Smith (1961, 1964).

Smith (1961) used the assigned purposes of getting a general impression and remembering details in reading two different parts of a biographical selection. She also gave readers a general question based on the content of the material and a suggested way of reading it. Twelfth grade students were used in this study. Differential test questions were used to measure the accomplishment of different assigned purposes. Smith also used tape recorded interviews of

introspective reports on the methods used for accomplishing each purpose and the past experiences of readers in reading for different purposes. In addition, she determined how well each subject held his assigned purpose in mind by asking each reader to state the purpose after he had read the assigned material. Although her results are not of primary interest, it is significant that she found that only approximately one-half of the poor readers (not defined) could remember the assigned purpose immediately after reading. This finding suggests that a question on the retention of purpose might well be used as a part of tests of reading flexibility using purpose as an independent variable.

Smith's experience in conducting the previous study probably led to the development of an informal Test of Purpose and an accompanying Reading Inventory for use by high school freshmen (Smith, 1964). The Test of Purpose provided an opportunity for students to select their own purposes as well as to read for assigned purposes. Part one of the test consists of twenty-four originally written selections of seventh-eighth grade readability about a variety of topics of interest to ninth grade students and written so as to be appropriate to a particular reading purpose. Passages range from 106-360 words in length. Students are asked to read each selection quickly and to choose the most appropriate purpose from a list of purposes.

Part two of the test consists of twelve selections prepared in the same manner as described above and designated to be read for a specific assigned purpose. The 12 assigned purposes were arrived at by starting with an original list of 215 purposes. Space does not allow a listing of these purposes, but in the writer's opinion, they constitute the most comprehensive and meaningful choice of purposes used in any test of reading flexibility. In addition, they are very specifically stated; e.g., "You are to read this selection for the purpose of understanding sensory images, or forming vivid images or pictures from a description (almost being able to see, hear, or touch objects)" (Smith, 1964). Five multiple-choice questions specifically designed to measure the accomplishment of the assigned purpose were provided for each passage. Each passage is timed to obtain a measure of rate. A checklist of reading approaches is used to obtain student introspections of procedures used in reading. Finally, the retention of each assigned purpose is determined following each article.

Two forms of the total test were constructed. Using seventy-three high school freshmen, the comparability of both forms was, pre-

sumably, attained by evidence of equal means and standard deviations on both parts of each test taken by the same students. The writer questions why correlations were not also used to study comparability. High reliability coefficients were obtained of about .90 for total test comprehension scores on both forms. It is strange that no attempt was made to demonstrate the reliability of the parts. They were probably low due to the use of only five test items. It is precisely the comparison of results on different parts of the test which provides the measure of reading flexibility. A crude attempt to validate the responses made to questions of procedures used in reading was made by using eye movement photographs and oral tape recorded retrospections. These results were, in some unexplained way, compared with initial test responses. It was claimed, without documentation, that similar responses in both situations gave evidence that students were in fact reading for different purposes.

The Reading Inventory to accompany the Test of Purpose consists of a checklist of the fifty-three statements of reading approaches. Students checked one of these columns to indicate if they usually, sometimes, or seldom read in the way described by the statement. A type of reliability check was made on a tentative draft of the test by interspersing throughout the inventory items meaning the same thing but worded differently. Results were not given. The writer wonders why a simple test-retest procedure was not used. In some undefined manner, eye movement photographs on the Test of Purpose were used to validate the inventory.

Despite some technical inadequacies and lack of norms and probably unreliable difference scores on rate and comprehension, this test should be used as a model for future test construction concerned with the role of purpose in reading flexibility. Also, unlike other tests which are supposed to measure changes in approach in relation to purpose, these tests were really designed to study the subtleties involved in the concept of reading approaches. The comprehensive and highly specific statements of purpose in the Test of Purpose are unsurpassed in tests or experimental investigations of reading flexibility. Due to the specificity of assigned purposes, the use of five test questions to measure the accomplishment of one purpose suggests the possibility for future testing of using criterion-referenced items which may be interpreted without reference to normative data.

Two interesting observations by Hill (1964) may have significant implications for the measurement of purpose-related reading flexi-

bility and the measurement of reading flexibility in general. Hill studied changes of rate and comprehension among college students reading three different articles for three different assigned purposes. All passages dealt with relatively complex materials on controversial social problems. He failed to obtain significant differences. Although he did not fully document this interpretation, he suggested that some evidence indicated individual interests in reading topics may have tended to affect the assigned purposes. A need to control the factor of interest has often been overlooked in the measurement of reading flexibility. Hill also allowed a smaller group to repeat the test a few days later. He obtained not only increases in rate and comprehension, as would be expected, but he also observed a great increase in reader interest. He raised the question of why the concept of reading flexibility should be restricted to single reading circumstances. Perhaps there is a suggestion here for a different measurement procedure involving more than one reading.

An experiment by Grant and Hall (1968) studied the relationship between comprehension and the reading achievement level of sixth grade students reading for two assigned purposes: 1) to read in order to answer a specific, broad, thought provoking question and 2) to read in order to answer questions. Important findings with suggestions for test construction were that, for the best readers who were reading at their independent level, there was no significant difference in comprehension; the average readers, reading at their instructional level, made significantly higher comprehension scores with the help of the thought provoking question; the poorest readers, reading at their frustration level, made slightly, although not significantly, lower scores on the passage without the thought provoking question. These results, like others, point to the importance of establishing the appropriate difficulty of material for the readers who will use a test designed to study the effects of other independent variables such as purpose. Observations by Henderson (1965), to be discussed, were in agreement with respect to the importance of choosing the correct difficulty of materials.

Most studies of the effects of purpose upon reader behavior have depended on assigned purposes, with the exception of Smith (1964). An experiment by Henderson (1965) investigated individually formulated purposes for reading among fifth grade students reading very easy material at the second grade level. Without discussing the details of this rather fully written study, it should be pointed out that a

significant relationship was found between ability to set one's own purpose and both general reading achievement and ability to attain a purpose. However, a more important finding from the standpoint of measurement techniques was that no significant differences in comprehension were found among the same subjects reading under three conditions: 1) purpose supplied by the experimenter, 2) development of own purpose after reading the first half of a story, and 3) no purpose assigned or designated by the student. Thus, it might follow that students who read well for assigned purposes may also read equally well when they choose their own purpose. More study, which might include rate as well as comprehension or other dependent variables, is needed on this question. It is not clear from reading this article whether there was any specific relationship between the assigned purposes and the comprehension test used. Also, the author noted that the reading material was written at the independent level for these subjects. He suggested that the relationship between skill and purpose setting and reading comprehension might increase with the use of more difficult material.

A method used by Bloomers and Lindquist (1944) has much to recommend it as a more adequate technique for allowing the reader to make a choice of rate and approach in order to attain an assigned purpose. This technique involves setting a specific purpose for each passage by providing a content-related question before each passage and directing readers, ". . . to read at a rate which seemed to them personally the most efficient for the accomplishment of the purpose set." This would appear to have some advantage over the usual admonition to read "as rapidly as possible."

As previously discussed in relation to Raygor's Reading Test, the techniques used in this test provide some freedom for each individual to choose his own purpose in the light of a brief description and the nature and difficulty of material.

More work is needed on the study of different ways of observing the effects of purpose on reading performance. In the writer's opinion, if a specific purpose is assigned on a test, some attempt should be made to ascertain whether or not the purpose was accepted or remembered. Also, more studies need to be carried out which allow students to determine their own purposes in testing situations.

Reader familiarity with material is another variable which needs to be controlled in the study of reading flexibility. So many studies point to the effect of previous knowledge upon scores of comprehension

tests taken after reading (Irion, 1925; Artley, 1944; Chall, 1947; Robinson, 1947; Dixon, 1951) that it is not necessary to describe studies on this obvious fact. More recently Preston (1962) has demonstrated that students answer questions beyond chance expectations without reading the passage upon which these tests were based, even on a standardized reading test.

If the effect of previous knowledge upon comprehension and/or rate test results is not controlled, then the differences between test performance on different passages cannot be interpreted as dependent variables which change as a function of variation of independent variables in the measurement of reading flexibility. In some story-type materials which are new to almost all readers, this factor may be of little consequence. Even with these types of materials, some readers more than others will have greater familiarity with the topic or style of the story. As the content of the test material becomes more closely related to specialized areas of experience such as subject matter fields, the likelihood of contamination of test results from previous knowledge increases.

With the exception of a test like the Braam-Sheldon Flexibility of Reading Test, familiarity of material is more important as a control variable than an independent variable in the measurement of reading flexibility. Current interest in constructing tests with reading-related items like the Carver-Darby Chunked Reading Test should speed up progress toward the development of reading tests less influenced by background knowledge than present tests.

There has been some speculation about different personality variables in relation to reading flexibility. However, people such as McDonald (1963) and Berg (1967) have been concerned with correlates of reading flexibility, such as psychological set or emotional freedom, which are not used as independent variables in the measurement of reading flexibility. Laycock (1958) studied the relationship of a personality variable to reading flexibility, but this study used personality as a correlate and not a variable in the measurement of reading flexibility.

Instructional set for reading rate has been used as an independent variable in reading flexibility measurement. Maxwell (1964) and Laycock (1955) found that college students could make significant improvements in reading rates, with adequate comprehension, by following simple instructions to read materials as fast as possible without loss of comprehension. These findings challenge the claim by

McDonald et al. (1968) that readers cannot change their reading rate at will. McDonald's point of view is based on a number of studies involving subjects of different ages using a variety of tests which involve a number of confounded variables. On the other hand Taylor (1960), in a well controlled study of the ability of eighth grade students to change reading rate under different instructional sets, found significant differences in rate but with a significant loss in comprehension. At present, the evidence about the influence of instructional set upon reading performance is conflicting. There is a need for replication of studies concerning the effects of the same instructional sets upon reading performance of readers of comparable ability using the same materials.

Bernstein (1955), Taylor (1960), and Bryant and Barry (1961) conducted studies on the influence of reader interest upon reader behaviors. Bernstein used two articles of equal difficulty, upon which significant differences in interest ratings had been determined for ninth graders, and an extensive comprehension test of thirty questions. She obtained a significant difference in comprehension in favor of the more interesting article. Bernstein's passages had been intentionally written to vary greatly in interest appeal. On the other hand Taylor (1960), using Reading Eye materials upon which significant differences in interest ratings had been found for college subjects, obtained no significant differences in eye movements, rate, or comprehension between passages. Similarly, Bryant and Barry (1961) found no significant differences in rate or comprehension between two articles read by two different groups of readers whose ratings indicated a perference for one article over the other. These articles were not very different in nature, and no indication was given that there was any extensive difference in interest value for these articles. In any case, such studies as these do not really measure reading flexibility. Neither losses nor increases in rate or comprehension based upon materials of various degrees of interest appeal necessarily indicate a desirable outcome of reader behavior in relation to changes in an independent variable. Interest appeal is an important control variable in the measurement of reading flexibility. There is no doubt that sufficiently great difference in interest appeal would affect various kinds of reader behavior.

A few other personality variables have been used as independent variables in relation to reader behaviors. Gifford and Marston(1966) studied the relationship between test anxiety, reading rates, and task

experience in an experimental study involving different groups of fourth grade boys. Half of the subjects received a practice test while the other half read the passage only once. Subjects read the passage for the purpose of getting main ideas or for remembering details. Differences were found in reading rate in favor of the subjects with low anxiety under no practice conditions. However, after practice, these differences disappeared. This significant finding is similar to the results of Rankin and Hess (1970) about the value of practice on reading performance of high-anxiety students. It also bears some relationship to Hill's suggestion (1964) about the value of rereading in relation to reading flexibility.

# Synthesis of Strengths and Weaknesses

## Areas of agreement and positive findings

### *Concepts of reading flexibility*

Although no attempt was made to review all of the definitions which have been formulated for "reading flexibility," it can be concluded that most definitions were in agreement that reading flexibility reflects the ability of a reader to change some aspects of his reading skills in order to attain a variety of reading purposes or to read different kinds of materials with good comprehension. The purposes assumed in most definitions involve different aspects of comprehension. Most definitions included some reference to the adjustment of rate in an efficient manner suitable to the reading task.

Various definitions were, therefore, in agreement that reading flexibility involves not merely changes in reading processes, but changes in a desirable direction in order to cope adequately with the demands of various reading tasks. It follows logically that measurements of reading flexibility necessarily involve observations of intraindividual changes in behavior in response to various circumstances.

The differences in definitions point to an important conclusion; namely, there is no one entity adequately described as reading flexibility. Instead, there are different kinds of reading flexibility. This conclusion has important implications for research, theory, and the development of measuring instruments.

McCracken's distinction (1965) between internal flexibility and external flexibility was an important one which has influenced subsequent development of a measurement procedure to measure internal flexibility.

### *Techniques of measuring reading flexibility*

*Published tests.* It is much easier to find differences than similarities among the few published tests of reading flexibility. All four tests

make provisions for the inducement of purposeful reading on all subparts. With the exception of the Flexibility of Reading Test by Braam and Sheldon, these tests attempt to have the reader read for several different comprehension related purposes. The purposes vary slightly, but they all provide an opportunity for the reader to read for thorough comprehension and to read for a more general understanding. Again, with the exception of the Braam and Sheldon test, all tests contain subtests which purport to measure the reader's ability to skim for main ideas and to scan in order to locate the answer(s) to one or more questions presented prior to reading the passage.

All tests, with the exception of the Test of Reading Flexibility by Spache and Berg, include materials of different degrees of readability, content, interest appeal, and novelty. These procedures reflect an omnibus general concept of reading flexibility which involves the effects of a large number of interacting variables upon reader behavior.

All of the published tests measure differences in reading rates obtained on different reading tasks. They also provide measures of comprehension for each task. Comprehension on these tests is measured by use of objective items which require a choice among alternatives. Most items are of the multiple-choice variety. Test items are used either in order to make it possible to interpret the rate scores or to measure the attainment of some purpose.

Each test provides some type of criteria to enable the reader to interpret his rate and comprehension scores. With the exception of the Test of Reading Flexibility by Spache and Berg, each test also provides criteria for the interpretation of differences in rates under various reading task conditions as an indication of reading flexibility.

Each test includes reading materials selected as suitable to the reading level and interests of the reader for whom the tests were written. The length of the passages are substantial on all tests.

All comparisons of rate and comprehension scores of the published tests of reading flexibility are based on data obtained after reading an entire passage rather than data obtained in the process of reading a passage. Thus, they all measure what McCracken (1965) terms *external flexibility*.

The Test of Reading Flexibility is the only published test which is designed to measure the effects of one independent variable (i.e., purpose) upon reading rate and comprehension. Other factors which might affect these reader behaviors are held constant while purpose is

systematically varied. This measurement design is highly commendable in that it provides results which can be easily interpreted. Another desirable feature of this test is that it does not attempt to use difference scores, as such, as criteria for reading flexibility. Instead, separate norms were provided for rate and comprehension. This is one means of avoiding the difficulties involved in obtaining suitable reliability on difference scores based on correlated measurements.

The Flexibility of Reading Test by Braam and Sheldon, although it possesses a number of technical deficiencies, nevertheless serves as a model for the measurement of a different kind of flexibility using subject matter as an independent variable.

The authors of both the Reading Versatility Tests and the Reading Test made commendable attempts to construct comprehension tests based on reading related items not subject to being answered by readers who had not read the passage.

The Reading Test by Raygor contains a desirable innovative feature in its skimming and scanning subtest. This subtest contains a variety of different materials such as indexes, charts, bibliographies, and textbook excerpts. No other published test uses this technique which closely resembles the types of materials that students would be likely to use in skimming and scanning in study-type reading. This subtest is also unique in providing a large number of items to measure the students' comprehension while skimming and scanning.

*Experimental measurements.* Most unpublished tests and measurement procedures used in research studies have attempted to measure the specific influence of variations in a particular causal agent, such as reading difficulty or purpose, upon changes in several reading behaviors. This type of test design has produced test results which can be interpreted with some precision. When an attempt was made to study the effects of more than one causal agent, the relative contribution of each factor to changes in behavior was measured.

A survey of the literature has revealed that experimental tests have studied the independent effects of many different causal agents upon many different reader behaviors. This is in great contrast to published reading tests. Examples of independent variables studied include the effects of content, style, familiarity with materials, instructional set for rate, interest, a wide variety of reading purposes and different ways of assigning them, the ability of students to determine their own purpose, and personality variables. Different reader behaviors studied include rate, general comprehension, eye-movements, introspective

reports of awareness of reading processes, retention of purpose, interest, and comprehension as attainment of a specific purpose.

There was considerable agreement in the literature that differences in the difficulty of material influence changes in a wide variety of reader behaviors (Judd and Buswell, 1922; Walker, 1938; Pitcher, 1953; Letson, 1959; Taylor, 1960; Levin, 1968).

There was general agreement in most studies that different purposes influence changes in various reader behaviors (Walker, 1938; Shores, 1960 and 1961; Smith, 1964; Grant and Hall, 1968; Levin, 1968).

Several studies have yielded results which emphasize the importance of the general level of reading difficulty used in tests of flexibility for readers at a given reading achievement level. There is some indication that material written at the reader's independent level is not suitable for measuring reading flexibility (Taylor, 1960; Henderson, 1965; Grant and Hall, 1968).

Results of several studies point to the conclusion that familiarity with materials influences reader behaviors (Irion, 1925; Artley, 1944; Chall, 1947; Robinson, 1947; Dixon, 1951; Preston, 1962).

Gifford and Marston (1966) and Rankin and Hess (1970) have found evidence that pretest practice (not just a few simple items as in most tests) prevents some test procedures from penalizing readers with a high test-measured anxiety level.

Several findings from individual studies suggest a number of important implications for future research and development of measurements of reading flexibility. It should be noted that replication of these studies is needed in order for great confidence to be placed in them.

Letson (1959) found that differences in material difficulty produced greater differences in reading rate than did differences in assigned purposes. He also observed that some readers have negative flexibility; that is, a tendency to vary rate in the opposite direction than would be expected from differences in material difficulty and purpose.

Pitcher (1953) found that differences in style interact with differences in reading difficulty in producing changes in reading rate.

Rankin and Hess (1970) have developed a new procedure for measuring intra-article reading flexibility that has potential as a research technique. Rankin (1970-1971), using this procedure, obtained evidence of greater reading flexibility, even among poor readers, than previous research on reading flexibility had indicated.

Smith's finding (1961) that only one-half of a group of poor readers among twelfth grade students could remember the assigned purpose on passages in a test immediately after reading a passage is of significance to test constructors. Her 1964 test contained some excellent techniques for studying reading approaches in relation to many different assigned purposes. This test provides an excellent model for the study of purpose.

Hill (1964) found evidence that differences in students' interest in materials had influence upon the effects of assigned purposes. He also obtained data that suggests that the concept of reading flexibility not be limited to materials read for the first time. He found that rereading a passage produced beneficial effects upon reading performance and interest.

Henderson's study (1965) suggested that readers who read well for assigned purposes will also read well when choosing their own purposes.

The study by Bloomers and Lindquist (1944) used a different way of stating an assigned purpose that has much to recommend it over the usual directions "to read as rapidly as possible in order to understand." The reader was told, after being presented with a content-related question, to read at a rate which seemed the most efficient for the successful accomplishment of the purpose.

The technique used by Raygor (1970) of describing the general nature and difficulty of selections and allowing students to choose their own purpose in the light of this information, is an innovative feature which might be emulated by other test constructors.

## Critique

### *Concepts of reading flexibility*

Various concepts of reading flexibility, as revealed in a variety of definitions, differ mainly with regard to the function of reading rate. Some authorities (Carrillo and Sheldon, 1952; Braam, 1963) have defined reading flexibility in such a way as to indicate that changes in rate per se, in relation to different purposes or materials, will bring about adequate comprehension. McDonald (1963, 1965, 1967) and others have defined the concept so as to regard rate changes as the result of utilizing different reading approaches in relation to various

purposes and material characteristics, so that adequate comprehension is attained. This dichotomy is confusing to researchers, test constructors, and teachers. It is quite understandable that different authorities should give different definitions of a concept. However, if this particular issue could be resolved through research stemming from well formulated theory or even logical analysis of known reading tasks, it would clear up a lot of confusion about the concept of reading flexibility.

In the writer's opinion, the chief deficiency in virtually all definitions of reading flexibility is the restriction of the meaning of purpose to aspects of comprehension. Even Stauffer (1962), who defined the concept as efficient and satisfactory attainment of purpose, proceeded to discuss examples which were restricted to types of comprehension. There are legitimate purposes for reading other than comprehension which might be included within a broader concept of reading flexibility.

### *Techniques of measuring reading flexibility*

Many of the deficiencies of published and experimental measurements are similar, hence no attempt is made in this section to organize the critique into these two subdivisions of published and experimental measurements. Instead, both types of measurement are discussed in relation to a given topic.

*Confounded variables.* In many tests of reading flexibility, the test is designed so that the differences in reader behavior cannot be interpreted as the result of variation in a single variable such as purpose or difficulty of material. Instead, changes in behavior as measured by the tests may be due to an unknown combination of factors such as interest in materials, style of materials, content of materials, difficulty of materials, familiarity with materials, and purpose. This lack of control makes flexibility scores difficult to interpret. Since two identical scores may have different meanings for two individuals, the test results are of little diagnostic value for teachers. The confounding of variables is more characteristic of published tests than of experimental measurements which often measure the independent and relative influence of two or more variables.

*Difference and ratio scores.* In most measurements, reading flexi-

bility is measured by differences in a reader's behavior (or ratios based upon differences) under two or more circumstances. This measurement procedure is apt to result in difference scores with low reliability unless the two measures are not highly correlated and each of the two measurements have high reliability coefficients. This problem is more acute for tests which were designed to measure individual differences than for experimental tests designed to measure group differences. Even in the latter case, difference scores with low reliabilities may result in findings of nonsignificant differences between group means. The writer has not found a single published test or experimental test which has given any evidence of sufficiently reliable difference scores.

The use of difference scores (or ratio scores) is also questionable on the grounds that a person whose lowest score is near the top of the distribution may not have as equal an opportunity to display a large difference score as a person whose lowest score is near the bottom of the distribution. Whether there is such measurement induced interaction is not known.

*Technical characteristics.* Manuals for published tests and research studies using experimental measurements have often provided either little information or questionable information about such things as reliability, validity, norms, comparability of subtest results, or comparability of all test forms.

Reliability coefficients are often missing (or lacking) for either rate or comprehension measurements in reading flexibility tests. Although it is not the writer's intention to carelessly or unjustly level accusations, several practices seem to falsely create the illusion of satisfactory reliability. For example, some published tests have combined two subtests in computing reliability, Some tests have inflated reliability coefficients by using Kuder-Richardson reliability coefficients for speeded tests. Some tests are accompanied by reliability figures for the total test but not for the subtests, which are crucial to the measurement of flexibility. One published test manual presents reliability coefficients for several tests, but not for all test forms in the battery. One published test is used despite a study which found that the reliabilities of its subtests were inadequate for both rate and comprehension.

Test validity is often ignored altogether. It is interesting that the writer has found not one study of concurrent validity using two different tests of reading flexibility. Often undocumented claims were made for content validity. One test used another published reading

test as a validity criterion for a flexibility of skimming and scanning subtest, even though the other test contained no measures of either set of skills. The test manuals and/or published studies about some tests contained some vague references to eye-movements as validity criteria, but presented no empirical evidence to show the specific results of this procedure.

Criteria for interpretation of rate, comprehension, or rate difference scores have been given often without any indication of how these criteria were obtained. Some criteria were apparently arbitrary. In some instances, norms were presented without any description of the normative sample.

Several tests have used comprehension subtests constructed to yield equal scores even though the tests were measures of performance on tasks of unequal difficulty. This practice raises serious questions about their validity. In general, then, most reading flexibility tests either do not meet the technical criteria for good test construction, or their constructors simply have provided no information about such matters.

*Comprehension measurement.* The confusion about the different functions of comprehension tests in the measurement of reading flexibility is discussed in relation to the writer's model of reading flexibility. However, it should be noted that comprehension tests were often used which are not adequate measures of the attainment of an assigned purpose. Another common fault is the construction of tests which are too short to have adequate reliability. The relationship of the comprehension measurement and the reader's rate is not clear in many tests. As Farr (1969), has indicated, the practice of forming a rate-of-comprehension index by multiplying rate times comprehension percentage is not justified. Some general comprehension tests allow the reader to refer to the article which answers questions, while others demand recall of information. Only a few test constructors have attempted to use reading related items which would be free from the influence of previous information on their comprehension tests. In general, there has been a tendency to construct comprehension tests which emphasize the measurement of lower level aspects of comprehension.

*Rate measurement.* Test constructors have used a variety of techniques in measuring rate. Some tests have used time limits in measuring rate while other tests have been untimed. Some tests have instructed the reader to read as rapidly as possible, while others have instructed the reader to read at his normal rate or at a rate suitable to

attaining the assigned purpose. One test measures rate based upon words read during the first three minutes, while most others allow a longer time. Some rate measurements were based upon the reading time before answering questions, while others included the time taken in answering question in computing rate. Such a variety of measurement practices call into question the comparability of different reading flexibility test results.

*Purpose.* Statements of purpose have, in most measurements, been limited to general types of comprehension components like remembering details or getting the main idea. Few studies have attempted to determine the extent to which assigned purposes are accepted. Purpose statements are many times vague and general and give the reader little help in adjusting reading approaches to purpose attainment. At present, it is not known whether it is best to assign specific purposes or to allow readers to choose their own purposes in the light of some information about the material. Some assigned purposes like "read to understand the article completely" probably serve to confuse the reader. Some statements of purpose include suggestions for attaining it, while others do not do this. Some assigned purposes are related to the content of the passages, while others bear no relationship to content. Such a variety of ways to assign purposes makes it difficult to compare the results of different reading flexibility tests. Indeed, interpretation of single tests are useless without careful definition of assigned purposes.

*Readability of materials.* The difficulty of materials used in a test of reading flexibility has a crucial bearing upon the results. Many published tests are not accompanied by sufficient information in the manual about this important factor. In contrast, most informal measurements do give the reader this information. The precise effects upon the measurement of reading flexibility of the readability level of materials in relation to the reading achievement level of the reader are not known. Even the effects of differences in readability between passages upon flexibility measurements are not known. The failure of some tests to find evidence of much reading flexibility in the population may be due to small differences in readability between different passages in the test.

*External flexibility.* Virtually all reading flexibility measurements of the effects of material difficulty upon reader behaviors have been limited to comparisons of results obtained on two or more complete passages. Very little is known about changes in adjustment of reading

approaches while a person is reading a single extended passage (internal flexibility).

*Negative flexibility.* The tendency of some people to slow down for easy passages or purposes and to speed up for more difficult passages or purposes has been observed in only one study. Little is known about this phenomenon.

*Effects of instructional set.* Conflicting evidence in the literature and competing claims of various authorities about the effects of instructions to read faster leave this issue undecided.

*Effects of interest.* Again, conflicting evidence on the effects of interest on reader behavior makes for uncertainty in test construction. It is not known how much difference in interest appeal two articles must have in order to produce a significant effect upon the reader's performance.

*Selection of dependent variables.* A wide variety of reader behaviors have been used in the measurement of reading flexibility. Such usage is related to the different concepts of flexibility previously discussed. It is not known, however, if a causal agent which produces a change in one reader response will produce similar changes in other reader responses.

*Generalization of findings.* There are many different ways of measuring reading flexibility which involve different pairings of independent and dependent variables, and which are based upon tests of unknown validity. These tests are taken by readers of different ages and reading levels, and are rarely replicated. Most generalizations about reading flexibility must, therefore, be interpreted with caution.

# Recommendations

The purpose of this concluding section is to make recommendations for future research and development on the measurement of reading flexibility, based upon the findings of this paper and a proposed reading model, which should result in clearer conceptions of the nature of reading flexibility(ies) and more valid and useful measurement instruments. First, a model is presented which is designated to clarify conceptualizations of different kinds of reading flexibility designated by the choice of the independent variable(s). Then designs are suggested for future research and development involving different control variables and dependent variables. After a discussion of the implications of the proposed model, a number of recommendations which stem from the review of literature are made for improvement of the measurement process.

## Model of reading flexibility

This model presents the general conception that reading flexibility involves an individual reader's ability to make desirable adjustments in reading approaches in order to enable him to attain one or more legitimate reading purposes under different conditions. These conditions are related to differences in materials, differences in his own psychological state, or differences in the external environment. It is assumed that the unobservable changes in approach can only be measured by studying observable changes in the reader's behavior in relation to changes in one or more independent variables. The term "legitimate reading purposes" designates a purpose which can reasonably be expected to be attained through reading.

## Implications of the model

Given these considerations, it follows that only intraindividual changes in behavior constitute valid dependent variables for the measurement

**Table 1 GENERALIZED MODEL FOR READING FLEXIBILITY**

| Independent (or control) Variables | Approach Variables | Reader Behavior Variables |
|---|---|---|
| *Materials* (Inter- or Intra-Passage)<br>*Readers* (Intra-Individual)<br>*External Environment* (Inter-Situation) | Perceptual, Cognitive, or Affective Processes Involved in Reading Purpose Attainment (Intra-Individual) | Rate, Comprehension, Retention, Evaluation, Application, Introspective Reports of Appreciation, Approach, or Purpose (Intra-Individual) |

of reading flexibility per se. Of course, all changes in reading approaches take place within the individual, and may, under suitable circumstances, result in behavioral changes. Independent variables include manipulated changes either between or within materials, within the reader himself, or between different environmental circumstances. These measured changes in behaviors may be observed, in the classical design, when only one independent variable is changed and all other relevant variables are held constant. Any given variable in the independent variable column may be, for one purpose, systematically changed as a causal agent or, for another purpose, serve as a control variable which is held constant. In multivariate designs, the effects of changes in several independent variables upon changes in reader behaviors can be studied with suitable statistical techniques. These observations, which have traditionally been made with reference to "experimental designs," have important implications for the measurement of reading flexibility. This is true even though precise control is lacking in the form of a test.

It is suggested that a more precise conception of reading flexibility would be attained if one or more verbal labels were added to the term "reading flexibility" based upon the specific independent variable(s) used in its measurement. For example, one might measure material

difficulty flexibility, purpose flexibility, noise level flexibility, or material difficulty plus purpose flexibility. When two or more independent variables are used, the measuring process should be designed to compare their differential effects upon reading behaviors. More precise designations based upon well-designed measurement procedures would improve conceptualizations of reading flexibility. Greater precision of measurement and communication about test results and research findings would result. Improvements should be circular and cumulative among concepts, measurement, design, and communication.

Examples of material variables might include readability, style, subject content, and typographical features. The writer would include under reader variables several features which have not been considered or measured as a type of reading flexibility. The concept of *purpose* might be expanded beyond conventionally measured comprehension purposes (understanding main ideas, remembering details) to include any legitimate purpose for reading, the outcomes of which could be measured. For example, reading for appreciation, reading for application, reading for memorizing a role in a play, reading for long range retention, and reading for evaluation, could all be used appropriately. Of course, as a measure of flexibility, the effects of two or more purposes would have to be compared. Previous knowledge about the content of a passage should, in most measurements, be held constant through passage selection as a controlled variable. Otherwise, the effects of differential reader knowledge should be taken into account in the measurement process. It should be possible, however, to vary this factor to determine the extent to which a reader can adjust his approach. It should be done in such a way as to minimize comprehension loss under conditions when background experience is lacking. Other reader variables, some of which might be used as either independent or control variables, are interest in topic, instructional set, fatigue, and mood. The last major category in the independent variable column is labeled "external environment." This, also, is usually supposed to be held constant while measuring other types of reading flexibility. However, the model would indicate the possibility of extending the concept of reading flexibility to include measuring changes in reader behavior under various environmental conditions such as noise level, presence of other people, and visual distractions. In summary, this model suggests both a more precise operational

definition of reading flexibility and an expansion of the concept to include a variety of independent variables not usually included.

A complete understanding of the many perceptual, cognitive, and affective components of reading approaches is lacking at present. These processes would include any and all internal processes which take place in response to a given reading task as a means of attaining the purpose of that task. Such processes as perceptual discriminations, perceptual clozure, synthesis, analysis, induction, drawing conclusions, comparisons, study skills, word attack skills, emotional responses, and motivational state would be involved. Only a relatively general idea of the variables involved in reading processes can, at present, be attained from behavioral evidence. The use of criterion referenced test items should shed light on this complex problem area.

Either one or several reader behavior variables might be used as empirical criteria for measuring the outcomes of changes in one or more independent variables. These might include such measured outcomes as rate, comprehension, eye-movements, long range retention, evaluation, application, introspective reports on appreciation, awareness of reading processes, or acceptance and/or retention of purpose, or ability to formulate an appropriate purpose. Generally, any measurable behavior may be used as a dependent variable provided that it is appropriate to a given concept of reading flexibility and is sensitive to changes in the independent variable(s) being used in the measuring process.

A note should be added about the role of comprehension as a product in flexibility measurement. One use of this variable is as a measure of purpose attainment. For example, if different assigned purposes included recall of details, interpreting author's intent, or drawing conclusions from evidence, then appropriate criterion referenced comprehension questions should be used to measure the reader's success in accomplishment of purpose. In other studies, comprehension questions might serve merely as a means of ascertaining the validity of rate scores. As such, these test results would not be, in effect, actual dependent variables as the term is used in the model. Still another use of comprehension test questions is to elicit in the reader a serious "mental set" which will increase the probability that he will cooperate in the measurement proceedings. Again, such a use of comprehension test results would not constitute a dependent variable in this model. The concept of reading flexibility stemming from the model is essentially positive in that it views reading

flexibility as a desirable outcome of learning. Therefore, comprehension *consistency* (at a level suitable to purpose) rather than *variation* in comprehension, despite changes in independent variables, could be an indication of reading flexibility.

This model suggests a general concept of reading flexibility in terms of a research design model. The model specifies the necessity for greater precision in both conceptualization and measurement procedures. It suggests an expansion of the concept which should stimulate new and different kinds of research and development of measurement procedures. As a research design model, it points to the kinds of control variables which should be utilized in measuring reading flexibility in the future.

## Other recommendations

This concluding section is based on evidence from both positive findings and critical evaluations of the literature on the measurement of reading flexibility. It is divided into two categories: 1) suggested procedures for improving measurements and 2) suggested research on measurement.

### *Suggested procedures for improving measurements*

*Confounded variables.* Tests of flexibility should be constructed so that the independent effects of separate independent variables upon reader behaviors can be determined. Tests constructed by Pitcher (1953), Letson (1959), and Levin (1968) could serve as basic models for such tests. One passage read for different purposes, as in the *Test of Reading Flexibility* by Spache and Berg, could serve as another model suitable to the attainment of this purpose. Results of these kinds of tests would be of diagnostic value to teachers, because the teacher would know what factor was responsible for a student's lack of reading flexibility.

*Difference and ratio scores.* Difference scores as measures of flexibility should not be used unless the difference between the two scores exceeds the standard error of measurement of a difference for the two tests. A formula for the standard error of measurement of the difference, denoted $S_{m\,Diff(1\text{-}2)}$, $= \sqrt{S^2_{m_1} + S^2_{m_2}}$, where $S_{m_1}$ is the

standard error of measurement for one test and $S_{m_2}$ is the standard error of measurement for the other (Thorndike and Hagen, 1961).

Future tests of reading flexibility should indicate evidence of the reliability of any difference scores used as measures of flexibility. The formula for this is: $r_{Diff} = \frac{\frac{r_{11} + r_{22}}{2} - r_{12}}{1 - r_{12}}$, where $r_{11}$ is the reliability of one measure, $r_{22}$ is the reliability of the other measure, and $r_{12}$ is the correlation between the two measures (Thorndike and Hagen, 1961). From this formula, it can easily be seen how the correlation between measures reduces the reliability of the difference scores.

Although it may not be apparent, the previously stated formula for the standard error of measurement of the difference does reflect the influence of the reliability of the difference, denoted $r_{Diff}$ since the $S_{mDiff(1-2)} = S_{Diff(1-2)} \sqrt{1-r_{Diff}}$.* Proof of this was provided by Dr. Frederick B. Davis of the University of Pennsylvania in personal correspondence.

The problems of interpreting ratio scores and measuring their reliability are so complex (Cronbach, 1941) that it might be best to substitute difference scores for ratios based upon differences.

The use of correlation coefficients as measures of flexibility as used by Rankin and Hess (1970) is one way to avoid the problems of difference scores. Another way of measuring flexibility without using difference scores is to provide separate norms for rate and comprehension (or other variables) as was done by the Test of Reading Flexibility by Spache and Berg.

*Technical characteristics.* Appropriate reliability coefficients should be provided for all measurements used in testing reading flexibility. Also, steps should be taken to improve reliability. The following suggestions by Jongsma (1971) should be helpful in attaining this objective: provide clear instructions, control external conditions, use long reading passages, use longer tests, and follow the technical criteria for test construction.

The validity of a reading flexibility test is not to be taken for granted. Evidence of validity such as content validity, concurrent

*The standard error of the difference, denoted $S_{Diff(1-2)}$, reflects the amount of difference between two scores attributed to differences in both "true score" and "error score." The standard error of the *measurement* of a difference, denoted $S_{mDiff(1-2)}$, reflects the amount of difference between scores attributed to differences in "error score."

validity, predictive validity, or construct validity should be provided for every measurement of reading flexibility.

Comparability of tests of comprehension based upon tasks of unequal difficulty should be accomplished not by trying to construct tests of equal difficulty, but by converting raw test scores to derived scores such as standard scores. Thus, tests results could be compared even though the raw scores are not equal in difficulty.

Adequate norms should be provided for all published reading flexibility tests with respect to rate, comprehension, flexibility measurement, or any other measurements used. Norms for difference scores might start with minimal differences which exceed their standard error and then provide percentile, or some other norms, to differences greater than this minimum.

*Comprehension measurement.* Constructors of reading flexibility tests should be familiar with the various functions of such tests in the measurement of reading flexibility. (See discussion in relation to the writer's Reading Flexibility Model.) Comprehension tests must suit the purposes for reading each passage. The use of criterion referenced items in certain types of flexibility measurements would be desirable and would remove the necessity of using norms for interpretation. Efforts should be made to construct tests with reading related items that could not be guessed by persons not reading the passages. Tests of comprehension should reflect many different aspects of comprehension, provided that this kind of test is suitable to the reader's purpose. Comprehension is best measured in power tests without time pressure.

*Rate measurement.* Rate measurements in reading flexibility tests should be based on a sufficiently long period of time to obtain a reliable measurement. With the exception of scanning tests, it is best not to confound rate and comprehension test measures by including the time taken to answer questions in the timing procedure. The directions given to the student related to rate of reading need to be appropriate to the purposes for reading. The validity of rate measurements must always be determined by the reader's comprehension.

*Purpose.* If purposes are assigned on a test, it would be wise to determine each student's acceptance of the assigned purpose through some type of introspective report. Statements of purpose must be precise and not subject to differing interpretations. The test by Smith (1964) should be used as a model for its excellent statement and selection of reading purposes. If the stated purposes do not help

readers to make decisions about how to adjust their reading procedures in order to accomplish the assigned purpose, they are of no value. In fact, they reduce the validity of the test. Also, the content of the passage should be suitable to its assigned purpose.

*Readability of materials.* Tentative research findings suggest the avoidance of easy materials written at the reader's independent level. It would probably be advisable to select reasonably large differences in readability of materials. Otherwise many of the resulting difference scores will not be reliable. The writer recommends the cloze procedure as an excellent measure of readability. Unlike readability formulae, it measures the readability of the passage for a particular group of readers.

*Effects of interest.* It is important to use materials of comparable interest appeal in most measurements of reading flexibility. This factor can be measured with rating scales or with the use of the semantic differential technique.

*Additional suggestions.* Future tests of scanning might use the skimming and scanning subtest of the Reading Test by Raygor as a model of good test construction. Finally, a valid test of flexibility using different subject materials similar to the Flexibility of Reading Test by Braam and Sheldon should be constructed. Such a test would be very helpful for teachers.

### *Suggested research on measurement*

*Confounded variables.* There is a need for replicated, well controlled studies on the specific effects of various independent variables upon different dependent variables involved in reading flexibility. Such studies should use multivariate designs to determine both the independent effects of each variable and also their interactions.

*Differences and ratio scores.* Research is needed on ways to improve the reliability of difference scores used on correlated measurements. Of course, the improvement of the reliability of each individual measurement would help. Also, the increase in differences in task difficulty to create greater differences in test performance might produce more difference scores which would exceed the magnitude of their standard error of measurement.

*Technical characteristics.* Research is needed on appropriate techniques for measuring the validity of reading flexibility tests. Little is known about this important problem and very little has been done in

determining the validity of flexibility tests, other than claims to the establishment of content validity, some comparisons of test results with eye-movement data, some correlations of comprehension tests with other comprehension tests, and comparison of test results with introspective reports. Techniques for measuring concurrent, predictive, and construct validity should be investigated.

*Comprehension measurement.* More research is needed which compares the practice of measuring comprehension by recall questions with the technique of allowing students to refer to the selection in answering questions. Study is also needed on techniques for constructing tests with reading dependent items. More knowledge should be obtained about the relationship of criterion referenced test items and classical test theory.

*Rate measurement.* More knowledge is needed on the formulation of instructions to students which influence rate of reading. What are the relative effects upon rate and comprehension of different instructions such as "read as rapidly as possible in order to answer questions," "read at your normal speed," "read at a rate which seems most efficient for attaining this purpose"?

Another matter in need of investigation is the extent to which changes in rate reflect changes in reading approaches suitable to different purposes. It is possible that, under certain circumstances, a change in rate might simply reflect an increase in speed of a particular reading process and not a change in approach. As Nacke (1971) has noted, we need to investigate precise ways of relating rate to comprehension. At present, the procedures for interpreting this relationship are vague.

*Purpose.* Study is needed on the relative effects of content related assigned purposes versus noncontent related assigned purposes. Research is also needed on the relative effects of including suggestions for purpose attainment together with a statement of purpose, as was done by Smith (1961), as opposed to using the statement of purpose without any such suggestions. More work is needed on the question of whether purposes should be assigned at all. Perhaps the reader should be free to determine his own purpose if given sufficient information upon which to make a decision. The latter condition constitutes the truest measure of reading flexibility under most realistic reading circumstances.

*Readability of materials.* More research is needed on selecting the readability of materials suitable for valid and reliable measures of different types of reading flexibility by various groups of readers.

Also, criteria for the selection of differences in the difficulty of materials for particular groups of readers need to be determined.

*Internal flexibility.* More investigation of the nature of internal (intra-article) flexibility and ways of measuring it is badly needed. A technological solution to the problem of machine scoring tests of internal flexibility as used by Rankin and Hess (1970) could be useful. Also, the relationship between internal and external flexibility is a problem worthy of investigation.

*Negative flexibility.* Only one investigator known to the writer has studied negative flexibility. More work is needed on ways of measuring this phenomenon and determining its significance.

*Effects of instructional set.* Conflicting evidence in the literature on this topic suggests the need for further research. This is a topic of practical significance to classroom teachers.

*Effects of interest.* Conflicting evidence on the effects of interest on reading performance—some of which is based on questionable research designs—suggests the need for more sophisticated study of this factor in relation to different age groups or students at various reading levels.

*Selection of dependent variables.* A vital issue to be determined by investigation is, "What kinds of reader behaviors are most suitable for the measurement of the effects of a given independent variable?" This question deserves systematic study. It is possible that the choice of one dependent variable might result in variations with a given independent variable while the choice of another dependent variable might not indicate any relationship at all.

*Other problems in need of solution.* Findings in this review suggest several other topics for future research: What are the effects of rereading a passage upon the measurement of reading flexibility? What are the effects of a practice test upon the measurement of reading flexibility?

*Generalization of findings.* With all of the limitations of current published tests and experimental procedures for measuring reading flexibility, generalizations are difficult to make. A program of systematic research and development along the lines suggested by the writer's model for reading flexibility, and the suggestions based upon this review of the literature, would greatly enhance the development of valid measurements of reading flexibility. With an organized approach based on the systematic investigation of hypotheses suggested by the model, research gaps could be eliminated and information useful to teachers could be obtained.

# References


**Artley, A. Sterl.** "A Study of Certain Relationships Existing between General Reading Comprehension and Reading Comprehension in a Specific Subject Matter Area," *Journal of Educational Research*, 37 (February 1944), 464-473.

**Berg, Paul C.** "Flexibility in Reading," in J. Allen Figurel (Ed.), *Vistas in Reading*, 1966 Proceedings, Volume 11, Part 1. Newark, Delaware: International Reading Association, 1967, 45-49.

**Berger, Allen.** "The Reliability of the Flexibility of Reading Tests," *Alberta Journal of Educational Research*, 15 (September 1969), 165-168.

**Bernstein, Margery R.** "Relationship between Interest and Reading Comprehension," *Journal of Educational Research*, 49 (December 1955), 283-288.

**Bloomers, P., and E.F. Lindquist.** "Rate of Comprehension of Reading: Its Measurement and Its Relation to Comprehension," *Journal of Educational Psychology*, 35 (November 1944), 449-473.

**Braam, Leonard S.** "Developing and Measuring Flexibility in Reading," *Reading Teacher*, 16 (January 1963), 247-251.

**Braam, Leonard S., and William D. Sheldon.** *Developing Efficient Reading.* New York: Oxford University Press, 1959, 298 pp.

**Brown, James I.** "The Flesch Formula 'Through the Looking Glass,' " *College English*, 13 (April 1952), 393-394.

**Bryant, N. Dale, and Nel E. Barry.** "The Relationships between Interest and Reading Rate and between Interest and Reading Comprehension," in Emery P. Bliesmer and Albert J. Kingston, Jr. (Eds.), *Phases of College and other Adult Reading Programs*, Yearbook of the National Reading Conference, 10 (1961), 127-130.

**Carlson, Thorsten R.** "Effects of Certain Test Factors on Measurement of Speed of Reading," *Journal of Educational Research*, 44 (March 1951), 543-549.

**Carrillo, Lawrence, and William D. Sheldon.** "The Flexibility of Reading Rate," *Journal of Educational Psychology*, 43 (May 1952), 299-305.

**Chall, Jeanne S.** "The Influence of Previous Knowledge on Reading Ability," Bulletin No. 26. Columbus: Ohio State University, 1947, 225-230.

**Cronbach, Lee J.** "The Reliability of Ratio Scores," *Educational and Psychological Measurement*, 1 (July 1941), 269-278.

**Dixon, W.R., W.E. Morse,** and **F.A. Ballantine.** *Studies in the Psychology of Reading.* University of Michigan Monographs in Education, No. 4. Ann Arbor: University of Michigan Press, 1951, 188 pp.

**Farr, Roger.** *Reading: What Can Be Measured?* Newark, Delaware: International Reading Association, 1969, 46.

**Flesch, Rudolph.** "New Facts about Readability," *College English*, 10 (January 1949), 225-226.

**Gifford, Edith M.,** and **Albert R. Marston.** "Test Anxiety, Reading Rate, and Task Experience," *Journal of Educational Research*, 59 (March 1966), 303-306.

**Grant, Eugene B.,** and **Marcia Hall.** "The Effect of a Thought-Directing Question on Reading Comprehension at Differing Levels of Difficulty," in J. Allen Figurel (Ed.), *Forging Ahead in Reading*, 1967 Proceedings, Volume 12, Part 1. Newark, Delaware: International Reading Association, 1968, 498-501.

**Gulliksen, H.** *Theory of Mental Tests.* New York: John Wiley and Sons, 1950, 486 pp.

**Henderson, E.H.** "A Study of Individually Formulated Purpose for Reading," *Journal of Educational Research*, 58 (July-August 1965), 438-441.

**Hill, W.R.** "Influence of Direction Upon the Reading Flexibility of Advanced College Readers," in Eric L. Thurston and Lawrence E. Hafner (Eds.), *New Concepts in College-Adult Reading*, Yearbook of the National Reading Conference, 13 (1964), 119-125.

**Humphrey, K.H.** "An Investigation of Amount Limit and Time Limit Methods of Measuring Rate of Reading," *Journal of Developmental Reading*, 1 (October 1957), 41-54.

**Irion, T.W.H.** *Comprehension Difficulties of Ninth Grade Students in the Study of Literature*, Contributions to Education, No. 189. New York: Bureau of Publications, Teachers College, Columbia University, 1925, 116 pp.

**Jongsma, Kathleen A.** "Reading Flexibility: An Investigation." Indiana University, 1971, 66 pp. (litho)

**Judd, Charles H.,** and **Guy T. Buswell.** *Silent Reading: A Study of the Various Types.* Chicago: University of Chicago, 1922, 157 pp.

**Laycock, Frank.** "Significant Characteristics of College Students with Varying Flexibility in Reading Rate," *Journal of Experimental Education*, 23 (June 1955), 311-330.

**Laycock, Frank.** "Flexibility in Reading Rate and Einstellung," *Perceptual and Motor Skills*, 8 (June 1958), 123-129.

**Letson, Charles T.** "The Relative Influence of Materials and Purpose on Reading Rates," *Journal of Educational Research*, 52 (February 1959), 238-240.

**Levin, Beatrice J.** "The Flexibility of Reading Rate," in J. Allen Figureal (Ed.), *Forging Ahead in Reading*, 1967 Proceedings, Volume 12, Part 1. Newark, Delaware: International Reading Association, 1968, 596-603.

**McCracken, R.A.** "Internal Versus External Flexibility of Reading Rate," *Journal of Reading*, 8 (January 1965), 208-209.

**McDonald, Arthur S.** "Factors Affecting Reading Test Performance," in Oscar S. Causey and Emery P. Bliesmer (Eds.), *Research and Evaluation in College Reading*, Yearbook of the National Reading Conference, 9 (1960), 28-35.

**McDonald, Arthur S.** "Flexibility in Reading," in J. Allen Figurel (Ed.), *Reading as an Intellectual Activity*, Proceedings of the International Reading Association, 8, 1963. New York: Scholastic Magazines, 81-85.

**McDonald, Arthur S.** "Research for the Classroom: Rate and Reading Flexibility," *Journal of Reading*, 8 (January 1965), 187-191.

**McDonald, Arthur S.** "Flexibility in Reading Approaches: Measurement and Development," in Mildred A. Dawson (Ed.), *Combining Research Results and Good Practice*, 1966 Proceedings, Volume II, Part 2. Newark, Delaware: International Reading Association, 1967, 67-71.

**McDonald, Arthur S.** et al. *Reading Versatility Tests* (Manual). Huntington, New York: Educational Developmental Laboratories, 1962 (revised 1968).

**Maxwell, Martha J.** "An Experimental Investigation of the Effect of Instructional Set and Information on Reading Rate," in Eric L. Thurston and Lawrence E. Hafner (Eds.), *The Philosophical and Sociological Bases of Reading*, Yearbook of the National Reading Conference, 14 (1964), 181-187.

**Maxwell, Martha J.** "Assessing Skimming and Scanning Skills Improvement," in George B. Schick and Merrill M. May (Eds.), *The Psychology of Reading Behavior*, Yearbook of the National Reading Conference, 18 (1969), 229-233.

**Nacke, Phil L.** "Assessment of Flexible Efficient Reading," in Frank P. Greene (Ed.), *Reading: The Right to Participate*, Yearbook of the National Reading Conference, 20 (1971), 256-265.

**Nicholaw, A. Constantine.** "Reading for Purpose: Instruction and Practice Effects on Rate, Comprehension, Flexibility in a Middle School Sixth Grade," unpublished doctoral dissertation, School of Education, Syracuse University, 1968.

**Pitcher, R.W.** "An Experimental Investigation of the Validity of the Flesch Readability Formula Related to Adult Materials," unpublished doctoral dissertation, University of Michigan, 1953.

**Preston, Ralph C.** "A New Approach to Judging the Validity of Reading Comprehension Tests: Summary of An Investigation," in J. Allen Figurel (Ed.), *Challenge and Experiment in Reading*, Proceedings of the International Reading Association, 7, 1962. New York: Scholastic Magazines, 166-167.

**Rankin, Earl F.** "How Flexibly Do We Read?" *Journal of Reading Behavior*, 1970-1971, 34-38.

**Rankin, Earl F., and Allen K. Hess.** "The Measurement of Internal (Intra-Article) Reading Flexibility," in George B. Schick and Merrill M. May (Eds.), *Reading: Process and Pedagogy*, Yearbook of the National Reading Conference, 19 (1970), 254-262.

**Raygor, Alton L.** *McGraw-Hill Basic Skills System: Reading Test* (Manual). New York: McGraw-Hill, 1970.

**Robinson, Francis P.** "The Effect of Language Style on Reading Performance," *Journal of Educational Psychology*, 38 (March 1947), 149-156.

**Shores, J. Harlan.** "Reading Science Materials for Two Distinct Purposes," *Elementary English*, 37 (December 1960), 546-552.

**Shores, J. Harlan** "Are Fast Readers the Best Readers? A Second Report," *Elementary English*, 38 (April 1961), 236-245.

**Smith, Helen K.** "Research in Reading for Different Purposes," in J. Allen Figurel (Ed.), *Changing Concepts of Reading Instruction*, Proceedings of the International Reading Association, 6 (1961). New York: Scholastic Magazines, 119-122.

**Smith, Helen K.** "The Development of Evaluation Instruments for Purposeful Reading," *Journal of Reading*, 8 (October 1964), 17-23.

**Spache, George.** "Diagnostic Tools," in Oscar S. Causey (Ed.), *Exploring the Goals of College Reading Programs*, Yearbook of the Southwest Reading Conference, 5 (1956), 121-123.

**Stauffer, Russell G.** "Speed Reading and Versatility," in J. Allen Figurel (Ed.), *Challenge and Experiment in Reading*, Proceedings of the International Reading Association, 7 (1962). New York: Scholastic Magazines, 206-210.

**Taylor, Stanford E., Helen Frackenpohl, and James L. Pettee.** "The Validity of Eye-Movement Photography as a Measurement of Reading Performance: Appendix Study No. II," *Eye Movement Photography With the Reading Eye*. Huntington, New York: Educational Development Laboratories, 1960, 7-12.

**Thorndike, R.L.,** and **E. Hagen.** *Measurement and Evaluation in Psychology and Education.* New York: John Wiley and Sons, 1961, 191-193.

**Traxler, Arthur E.** "The Relationship Between the Length and the Reliability of a Test of Rate of Reading," *Journal of Educational Research,* 32 (September 1938), 1-2.

**Troxel, Vernon E.** "Reading Eighth Grade Mathematical Materials for Selected Purposes," unpublished doctoral dissertation, University of Illinois, 1959.

**Walker, Robert Y.** "A Qualitative Study of the Eye Movements of Good Readers," *American Journal of Psychology,* 50 (July 1938), 472-481.

**Welch, Alice H.** "A Study of Eye Movement Patterns to Determine the Relationship Between Efficiency and Flexibility in Reading," unpublished manuscript, Marquette University Library, 1964.

# Test References

**Carver-Darby Chunked Reading Test.** Ronald P. Carver and Charles A. Darby, Jr. Silver Spring, Maryland: American Institutes for Research, 1970.

**Cooperative English Tests: Reading Comprehension.** Princeton, New Jersey: Educational Testing Service, 1960.

**Davis Reading Test.** Frederick B. Davis and Charlotte C. Davis. New York: Psychological Corporation, 1958 (revised 1962).

**Diagnostic Reading Tests.** New York: Committee on Diagnostic Reading Tests, 1947 (revised 1963).

**Flexibility of Reading Test.** Leonard S. Braam and William D. Sheldon, *Developing Efficient Reading.* New York: Oxford University Press, 1959, 120 pp.

**McGraw-Hall Basic Skills System: Reading.** Alton L. Raygor. New York: McGraw-Hill, 1970.

**Nelson-Denny Reading Test.** M.J. Nelson and E.C. Denny. Boston: Houghton Mifflin, 1960 (revised 1965).

**Reading Versatility Tests.** Arthur S. McDonald et al. Huntington, New York: Educational Developmental Laboratories, 1962 (revised 1968).

**SA-S Senior Scales.** Donald E.P. Smith, A.L. Wood, and P. Carrigan. Ann Arbor: Bureau of Psychological Services, University of Michigan, 1957.

**Test of Reading Flexibility.** George D. Spache and Paul C. Berg, *Faster Reading for Business.* New York: Thomas Y. Crowell Company, 1958, 298 pp.